

# Semantic Breakthrough in Drug Discovery

# Synthesis Lectures on the Semantic Web: Theory and Technology

Editors

**Ying Ding**, *Indiana University*

**Paul Groth**, *VU University Amsterdam*

Founding Editor Emeritus

**James Hendler**, *Rensselaer Polytechnic Institute*

Synthesis Lectures on the Semantic Web: Theory and Application is edited by Ying Ding of Indiana University and Paul Groth of VU University Amsterdam. Whether you call it the Semantic Web, Linked Data, or Web 3.0, a new generation of Web technologies is offering major advances in the evolution of the World Wide Web. As the first generation of this technology transitions out of the laboratory, new research is exploring how the growing Web of Data will change our world. While topics such as ontology-building and logics remain vital, new areas such as the use of semantics in Web search, the linking and use of open data on the Web, and future applications that will be supported by these technologies are becoming important research areas in their own right. Whether they be scientists, engineers or practitioners, Web users increasingly need to understand not just the new technologies of the Semantic Web, but to understand the principles by which those technologies work, and the best practices for assembling systems that integrate the different languages, resources, and functionalities that will be important in keeping the Web the rapidly expanding, and constantly changing, information space that has changed our lives.

Topics to be included:

- Semantic Web Principles from linked-data to ontology design
- Key Semantic Web technologies and algorithms
- Semantic Search and language technologies
- The Emerging "Web of Data" and its use in industry, government and university applications
- Trust, Social networking and collaboration technologies for the Semantic Web
- The economics of Semantic Web application adoption and use
- Publishing and Science on the Semantic Web
- Semantic Web in health care and life sciences

Semantic Breakthrough in Drug Discovery  
Bin Chen, Huijun Wang, Ying Ding, and David Wild  
2014

Semantics in Mobile Sensing  
Zhixian Yan and Dipanjan Chakraborty  
2014

Provenance: An Introduction to PROV  
Luc Moreau and Paul Groth  
2013

Resource-Oriented Architecture Patterns for Webs of Data  
Brian Sletten  
2013

Aaron Swartz's A Programmable Web: An Unfinished Work  
Aaron Swartz  
2013

Incentive-Centric Semantic Web Application Engineering  
Elena Simperl, Roberta Cuel, and Martin Stein  
2013

Publishing and Using Cultural Heritage Linked Data on the Semantic Web  
Eero Hyvönen  
2012

VIVO: A Semantic Approach to Scholarly Networking and Discovery  
Katy Börner, Michael Conlon, Jon Corson-Rikert, and Ying Ding  
2012

Linked Data: Evolving the Web into a Global Data Space  
Tom Heath and Christian Bizer  
2011

© Springer Nature Switzerland AG 2022

Reprint of original edition © Morgan & Claypool 2015

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Semantic Breakthrough in Drug Discovery

Bin Chen, Huijun Wang, Ying Ding, and David Wild

ISBN: 978-3-031-79455-1 paperback

ISBN: 978-3-031-79456-8 ebook

DOI 10.2200/978-3-031-79456-8

A Publication in the Morgan & Claypool Publishers series

*SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND TECHNOLOGY*

Lecture #9

Series Editors: Ying Ding, *Indiana University*

Paul Groth, *VU University Amsterdam*

Founding Editor Emeritus: James Hendler, *Rensselaer Polytechnic Institute*

Series ISSN

Print 2160-4711 Electronic 2160-472X

# Semantic Breakthrough in Drug Discovery

Bin Chen  
Stanford University

Huijun Wang  
Merck

Ying Ding  
Indiana University

David Wild  
Indiana University

*SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND  
TECHNOLOGY #9*

## ABSTRACT

The current drug development paradigm—sometimes expressed as, “One disease, one target, one drug”—is under question, as relatively few drugs have reached the market in the last two decades. Meanwhile, the research focus of drug discovery is being placed on the study of drug action on biological systems as a whole, rather than on individual components of such systems. The vast amount of biological information about genes and proteins and their modulation by small molecules is pushing drug discovery to its next critical steps, involving the integration of chemical knowledge with these biological databases. Systematic integration of these heterogeneous datasets and the provision of algorithms to mine the integrated datasets would enable investigation of the complex mechanisms of drug action; however, traditional approaches face challenges in the representation and integration of multi-scale datasets, and in the discovery of underlying knowledge in the integrated datasets. The Semantic Web, envisioned to enable machines to understand and respond to complex human requests and to retrieve relevant, yet distributed, data, has the potential to trigger system-level chemical-biological innovations. Chem2Bio2RDF is presented as an example of utilizing Semantic Web technologies to enable intelligent analyses for drug discovery.

## KEYWORDS

drug discovery, semantic data integration, semantic analytics, semantic graph mining, semantic prediction

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background	1
1.2	Data Representation in the Semantic Web	4
1.3	Data Query, Management, and Integration	7
1.4	Knowledge Discovery in Semantically Integrated Datasets	11
1.5	Chem2Bio2RDF	16
<b>2</b>	<b>Data Representation and Integration Using RDF</b>	<b>19</b>
2.1	Background	19
2.2	Methods	20
2.3	Discussion	31
2.4	Conclusion	34
<b>3</b>	<b>Data Representation and Integration Using OWL</b>	<b>35</b>
3.1	Introduction	35
3.2	System and Methods	36
3.3	Implementation	41
3.4	Use Cases	43
3.5	Discussion	48
3.6	Conclusion	48
<b>4</b>	<b>Finding Complex Biological Relationships in PubMed Articles using Bio-LDA</b>	<b>49</b>
4.1	Introduction	49
4.2	Materials and Methods	51
4.2.1	Databases	51
4.2.2	Bio-LDA	54
4.3	Experimental Results	56
4.3.1	Analyzing the Bio-LDA Model Results	56
4.3.2	Comparing the Bio-LDA and LDA Models	58
4.3.3	Identification of Bio-term Relationships within Topics	60

4.3.4	Discovery of Bio-term Associations .....	62
4.4	Application Tools .....	64
4.4.1	Literature Association Score Calculator (LASC) .....	64
4.4.2	Associated Bio-Terms Finder (ABTF) .....	67
4.5	Conclusion .....	70
<b>5</b>	<b>Integrated Semantic Approach for Systems Chemical Biology Knowledge</b>	
	<b>Discovery .....</b>	<b>71</b>
5.1	Introduction .....	71
5.2	Datasets .....	72
5.3	Methods .....	72
5.3.1	Association Prediction .....	73
5.3.2	Association Search .....	74
5.3.3	Association Exploration .....	75
5.3.4	Connectivity-map Generation .....	75
5.3.5	Chem2Bio2RDF Extension .....	76
5.4	Application Tools .....	76
5.4.1	Association Predictor .....	76
5.4.2	Association Searcher .....	78
5.4.3	Association Explorer .....	79
5.5	Use Cases .....	80
5.5.1	Identifying Potential Drugs for a Target .....	80
5.5.2	Investigating Drug Polypharmacology Using Association Search .....	82
5.5.3	Building a Disease-Specific Drug-Protein Connectivity Map .....	86
5.5.4	Association Search for Discovery Compounds .....	86
5.6	Conclusion .....	89
<b>6</b>	<b>Semantic Link Association Prediction .....</b>	<b>91</b>
6.1	Introduction .....	91
6.2	Materials and Methods .....	92
6.2.1	Network Building .....	92
6.2.2	Drug Target Pair Preparation .....	95
6.2.3	Path Finding .....	95
6.2.4	Statistical Model .....	95
6.2.5	Model Evaluation .....	96
6.2.6	Assess Drug Similarity .....	97
6.3	Results .....	98



6.3.1	Semantic Linked Data .....	98
6.3.2	Pattern Score Distribution .....	99
6.3.3	Pattern Importance .....	99
6.3.4	Association Scores of Drug Target Pairs .....	100
6.3.5	Comparison with Connectivity Maps.....	103
6.3.6	Assessing Drug Similarity from Biological Function .....	104
6.4	Web Services .....	105
6.5	Discussion.....	108
<b>7</b>	<b>Conclusions .....</b>	<b>111</b>
	<b>References .....</b>	<b>113</b>
	<b>Authors' Biographies .....</b>	<b>131</b>